

The International Journal of Digital Curation

Volume 7, Issue 1 | 2012

Developments in Research Funder Data Policy

Sarah Jones,
Digital Curation Centre,
University of Glasgow

Abstract

This paper reviews developments in funders' data management and sharing policies, and explores the extent to which they have affected practice. The Digital Curation Centre has been monitoring UK research funders' data policies since 2008.¹ There have been significant developments in subsequent years, most notably the joint Research Councils UK's *Common Principles on Data Policy* and the Engineering and Physical Sciences Research Council's *Policy Framework on Research Data*. This paper charts these changes and highlights shifting emphasises in the policies. Institutional data policies and infrastructure are increasingly being developed as a result of these changes. While action is clearly being taken, questions remain about whether the changes are affecting practice on the ground.

¹ Digital Curation Centre policy webpages: <http://www.dcc.ac.uk/resources/policy-and-legal/funders-data-policies>

Building Momentum in Funder Data Policy

There have been two major peaks in the development of research funder data policy: the first in 2007 and the second in 2010-11. The timeline in Figure 1 tracks the development of funders' data policies and the overarching requirements that have underpinned and driven change.

Unsurprisingly, the first UK research funders with data policies were those that also supported data centres: the Natural Environment Research Council (NERC), the Economic and Social Research Council (ESRC) and the Arts and Humanities Research Council (AHRC). NERC issued its first data policy handbook in 1996, and released updated versions in 1999 and 2002. This covered planning for data management, the responsibilities of NERC and data holders, and arrangements for access to data. In recent years, NERC has revised its data policy significantly, releasing a new policy last year (NERC, [2010](#)). The ESRC has had a similar trajectory in terms of policy development, with an early data policy covering the acquisition, maintenance and support of datasets in place in 2000. The Council subsequently provided a datasets policy as an annex in its Research Funding Guide, and now directs researchers to its new *Research Data Policy* (ESRC, [2010](#)). While the AHRC has not released a formal data policy, it has issued requirements related to data access and preservation since the late 1990s. These are provided in the *Research Funding Guide* (AHRC, [2011](#)) under sections on the 'Technical Appendix' and 'deposit of resources and datasets.'

A key driver in the development of further data policies came in 2004, when the UK joined the governments of 33 other countries in adopting the OECD *Declaration on Access to Data from Public Funding*. Principles and Guidelines were subsequently released to help guide the development of policies and good practices related to the accessibility, use and management of research data (OECD, [2007](#)). UK research funders responded to this impetus. The Medical Research Council (MRC, [2011](#)), Biological and Biomedical Sciences Research Council (BBSRC, [2010](#)), and Wellcome Trust ([2011](#)) all introduced data policies in 2005-7, which reference the OECD Principles as a starting point.

Latterly, we have seen a push towards harmonisation in funders' data policy, with the Research Councils UK (RCUK) *Common Principles on Data Policy* being released in April 2011 (RCUK, [2011a](#)). The Engineering and Physical Sciences Research Council (EPSRC, [2011](#)) and Science and Technology Facilities Council (STFC, [2011](#)) released their own data policies shortly thereafter, in May and September respectively. Disciplinary consortia have also united in terms of policy statements. The declaration on *Sharing Research Data to Improve Public Health* was signed by 17 major international public health research funders in January 2011 (Wellcome Trust, [2011](#)). Such consistency and coherence strengthens the requirements.

Research data policy developments

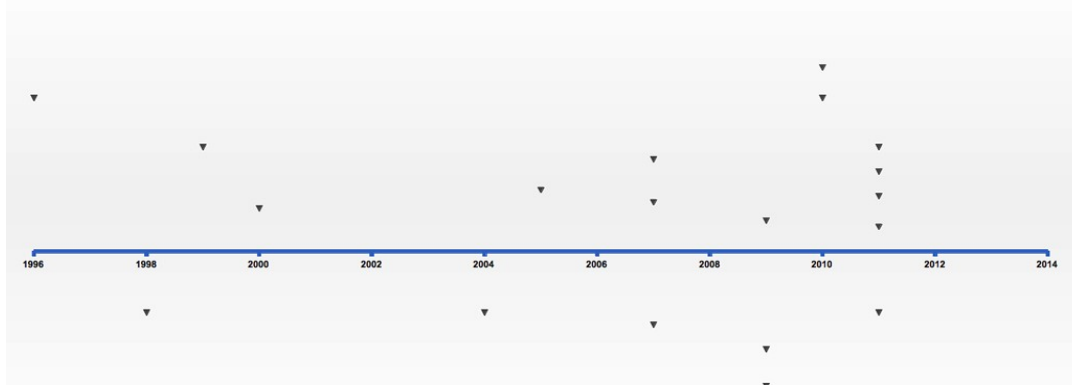


Figure 1. Research data policy developments timeline © 2011 Digital Curation Centre.

Policy	Date
NERC Data Policy Handbook	1996
RCUK Safeguarding Good Scientific Practice	December 1998
AHRC Dataset Requirements	c.1999
ESRC Data Policy	April 2000
OECD Declaration on Access to Research Data from Public Funding	January 2004
MRC Policy on Data Sharing and Preservation	2005
OECD Principles and Guidelines for Access to Research Data from Public Funding	2007
Wellcome Trust Policy on Data Management and Sharing	January 2007
BBSRC Data Sharing Policy	April 2007
Cancer Research UK Policy on Data Sharing and Preservation	July 2009
RCUK Code of Conduct and Policy on the Governance of Good Research Conduct: Integrity, Clarity and Good Management	July 2009
UKRIO Code of Practice for Research: Promoting Good Practice and Preventing Misconduct	September 2009
ESRC Research Data Policy	September 2010
NERC Data Policy	September 2010
RCUK Common Principles on Data Policy	April 2011
EPSRC Policy Framework on Research Data	May 2011
University of Edinburgh Research Data Management Policy	May 2011
STFC Scientific Data Policy	September 2011

Table 1. Timeline of research data policies from 1996-2011.

Overarching governance has been provided throughout this period through good research practice codes. A joint statement on *Safeguarding Good Scientific Practice* was issued by the Director General of the Research Councils and the Chief Executives of the UK Research Councils in December 1998. This highlights the central role of data, stating that:

“Primary data as the basis for publications should be securely stored for an appropriate time in a durable form under the control of the institution of their origin.” (RCUK, [1998](#))

Many universities introduced similar codes of good research practice in the early 2000s. However, despite these policies being in place, only a handful of institutions have made significant headway in terms of addressing research data management. In 2009 both the RCUK’s *Policy and Code of Conduct on the Governance of Good Research Conduct* (RCUK, [2011b](#)) and the UK Research Integrity Office’s *Code of Practice for Research* (UKRIO, [2009](#)) were released, putting increasing pressure on institutions to respond accordingly. Adopting the UKRIO code was noted as a prime driver in the development of the University of Edinburgh’s Research Data Policy (Rice & Haywood, [2011](#)).

The basic requirement to manage and provide access to research data has been in force for some time, yet a lack of budget and blurred responsibilities has allowed this to be deflected. In recent years momentum has built and data policies have become more co-ordinated and exacting. Policy statements have also become more enabling, providing practical guidelines and financial support which allow greater scope for implementation.

Harmonisation of Policy: The RCUK Common Principles

The RCUK’s *Common Principles on Data Policy* represents a significant step in the open data movement, taking the OECD Principles as their steer. Access and reuse is the common thread that unites all seven principles, as demonstrated in the table below. Data management and preservation are very much a means to an end. The ultimate goal is ensuring access and reuse of data of long-term value.

RCUK Principles	Key Message
Publicly funded research data are a public good, produced in the public interest, which should be made openly available with as few restrictions as possible in a timely and responsible manner that does not harm intellectual property.	Open access
Institutional and project specific data management policies and plans should be in accordance with relevant standards and community best practice. Data with acknowledged long-term value should be preserved and remain accessible and usable for future research.	Preservation for continued access

RCUK Principles (continued)	Key Message
To enable research data to be discoverable and effectively re-used by others, sufficient metadata should be recorded and made openly available to enable other researchers to understand the research and re-use potential of the data. Published results should always include information on how to access the supporting data.	Open metadata to support access/reuse
RCUK recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these are considered at all stages in the research process.	Legally/ethically appropriate release of data for reuse
To ensure that research teams get appropriate recognition for the effort involved in collecting and analysing data, those who undertake Research Council funded work may be entitled to a limited period of privileged use of the data they have collected to enable them to publish the results of their research. The length of this period varies by research discipline and, where appropriate, is discussed further in the published policies of individual Research Councils.	Embargo periods for privileged use
In order to recognise the intellectual contributions of researchers who generate, preserve and share key research datasets, all users of research data should acknowledge the sources of their data and abide by the terms and conditions under which they are accessed.	Acknowledge data sources and respect access conditions
It is appropriate to use public funds to support the management and sharing of publicly-funded research data. To maximise the research benefit which can be gained from limited budgets, the mechanisms for these activities should be both efficient and cost-effective in the use of public funds.	Cost-effective data management and sharing

Table 2. Identifying access and reuse as the common thread in the RCUK Common Principles on Data Policy

There are a number of significant points in the Common Principles which should help to advance data management and sharing practice. The second principle assumes that institutions and projects have data management policies and plans in place, adding further weight to the calls to develop these. Meanwhile, the third principle expects metadata to be made openly available to ensure that research data are discoverable and can be reused. Importantly, this principle also requires that published results include information on how to access the supporting data. These practical suggestions should provide more recognition for data. Perhaps most crucially, the

final principle confirms that it is appropriate to use public funds to support the management and sharing of publicly-funded research data, giving institutions more opportunities to finance this.

Naturally, it will take several years for these Principles to transform into common practice, but in collectively stating them, RCUK has made significant headway. The Principles will enable the development of new procedures, structures and guidelines to facilitate change. Some universities, for example, are investigating the use of existing research office systems or their Institutional Repositories to record and share metadata on their research data holdings. Citation mechanisms are being developed, and as more publications link to underlying data it is likely that standardised approaches and publisher conventions for doing this will emerge. Guidance has also been requested to help institutions cost data management and sharing. Associated costs are rarely built into grant applications and there remains a lack of clarity as to how these activities should be allocated in practice. The next few years will provide a number of useful test cases, as the Principles start to be adopted.

Shifting Emphasis in Funder Data Policy

A Focus on Access and Data Sharing

The RCUK *Common Principles on Data Policy* are indicative of wider change. Many individual research funder policies similarly emphasise access and data sharing. Preservation requirements are typically limited to a general statement that data should be maintained and remain accessible for ten years, whereas data sharing stipulations are far more exacting. Timeframes for the release of data are common, mechanisms for sharing and places of deposit are often suggested, and firm upper limits tend to be set for any restrictions, such as embargo periods. In terms of data management and sharing plans, a similar weighting towards data sharing is evident. The themes and questions funders propose seek to tease out precise details about when, how and to whom data will be shared, whereas practical details about data management, such as storage and back-up procedures, are only requested in a couple of cases.

More Pragmatic and Enabling Statements

Increasingly the content of policies is advancing beyond general principles to give more practical actions that can be implemented. Several funders, for example, ask for published results to include information on how to access the supporting data. The EPSRC policy, in particular, outlines a number of precise expectations, such as publishing metadata within 12 months, providing robust Digital Object Identifiers and not storing data in a jurisdiction with lower legal safeguards than the UK. A fair degree of pragmatism is also evident in the policies. NERC and the Wellcome Trust have introduced the notion of data value to help define what should be preserved, and most funders advise restricting data management and sharing to appropriate cases where there is clear scientific benefit and cost effectiveness.

Clearer Roles and Responsibilities

The 2009 RCUK *Policy and Code of Conduct* notes that data management is a shared responsibility between researchers and research organisations. However, specific responsibilities are often poorly defined across the various roles, resulting in confusion about how undertake data management in practice. Some subsequent policies have offered more guidance. The 2010 ESRC policy provides detailed implementation guidance, stating the specific responsibilities of grant applicants, grant holders, the ESRC and its data service providers. Funders' policies tend to place the onus on researchers to consider and make provision for research data management, largely through outlining their ideas in a data management and sharing plan. In contrast, the recent EPSRC policy places the responsibility squarely at the institution's door, which has caused a considerable stir.

The Significance of the EPSRC Policy Framework

The EPSRC policy is more exacting than that of the other UK research councils, providing very detailed expectations for metadata coverage and a preservation requirement that extends with each third-party access request. The scale of the challenge for institutions to maintain an accurate register of all their research data and related access requests cannot be underestimated.

Most notably, the expectations are placed on research organisations in receipt of EPSRC funding, rather than individual grant holders. Universities are better placed than individuals to provide infrastructure and implement systems for data management and sharing. The EPSRC outlines very clear timescales for this to be achieved: it expects all those it funds to have developed a clear roadmap to align their policies and processes with its expectations by 1st May 2012, and to be fully compliant with these expectations by 1st May 2015.

The significance of this move becomes apparent when you look at the income research-led institutions derive from the EPSRC. Comparing the value of research grants awarded by RCUK members in 2009-2010 shows that the EPSRC allocated the largest amount: £530m. EPSRC data on the value of awards for all current projects on 1st November 2010, brings into sharper focus the level of support provided to research-led Universities. Russell Group universities dominate the top end of the scale, with most holding grants to the value of several hundred million. Indeed, the average value of current EPSRC grants per organisation stood at £151m on the audit date.

Institutions are typically more risk-averse than individuals and given the proportion of grant income many derive from EPSRC, the risk of losing this is significant. The ambitious timescales EPSRC has set, and its commitment to monitor progress and implement appropriate sanctions, has understandably moved many HEIs to act. Seven out of 17 (41%) projects funded under the infrastructure strand of the 2001-13 JISC Managing Research Data programme noted the EPSRC policy as a driver in their grant applications, and many of those receiving tailored research data management support from the DCC are focused on developing a roadmap to comply with the EPSRC expectations.

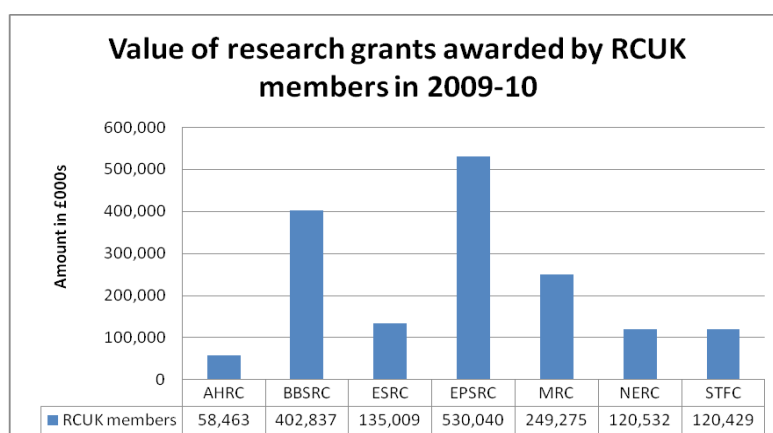


Figure 2. Value of research grants awarded by RCUK members in 2009-10. © 2011 DCC.²

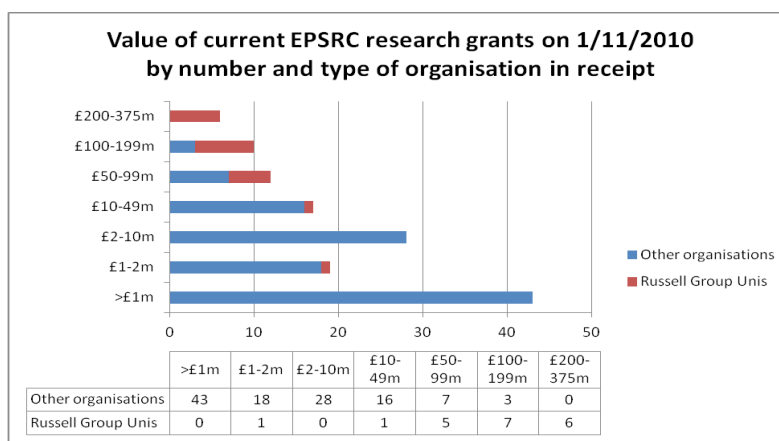


Figure 3. Value of current EPSRC research grants on 1st November 2010 by the number and type of organisations in receipt of awards.³ © 2011 DCC.

Institutional Responses to Implement Data Policy

Policies cannot be implemented without significant investment for services and infrastructure. Opportunities for institutions to respond to policy requirements are on the increase as various funds emerge. The research councils have jointly endorsed the use of public funds to support the management and sharing of publicly-funded research data, JISC has made a significant investment in research data management projects, and HEFCE has committed several million pounds to shared services. Mandates are finally being matched with money.

² Figures are taken from the financial statements in the 2010 Annual Report of each Research Council.

³ Figures from EPSRC Grants on the Web post-November 2010 grants portfolio at: <http://gow.epsrc.ac.uk/NGBOListOrganisations.aspx>. Note that the data represents all grants that were current on 1st November 2010, including announced grants that had not yet started and completed grants where financial reconciliation was still to take place. Some of the funds provided may cover access to facilities and have not been paid direct to the institution.

Many UK universities are now undertaking some work in the area of research data management and the results are inspiring. Three of the most advanced are the Universities of Edinburgh, Oxford and Southampton, each of which has undertaken several projects in recent years to develop institutional infrastructure and services.

A paper presented by Robin Rice at the 6th International Digital Curation Conference in 2010 outlined research data management initiatives at the University of Edinburgh (Rice, 2011). The DISC-UK DataShare project, run in collaboration with the Universities of Oxford and Southampton, allowed them to establish an institutional data repository and work with researchers to encourage data sharing. Parallel activity through the Data Audit Framework project investigated researchers' data management practices and needs for support. The findings and recommendations from this work have informed the development of various policies and strategies led by Information Services. The University released an exemplary *Research Data Management Policy* (University of Edinburgh, 2011) and has also been developing practical training and guidance for researchers, most notably through the Research Data MANTRA project.

The University of Oxford⁴ was involved in the DataShare project and was also a pathfinder for the UK Research Data Service. Two consecutive JISC projects scoped out digital repository services for research data management and embedded institutional data curation services in research. The latter project – EIDCSR – involved collaboration with the University of Melbourne on research data policy and led to a suite of data management guidance webpages. Projects funded under the 2009-2011 JISC Managing Research Data (MRD) programme addressed the specific needs of researchers in the humanities and life sciences. Key outputs from both of these projects are being developed as shared services: the Database as a Service model developed in Sudamih is now being rolled out in the ViDaaS project and the Admiral data management infrastructure is being progressed as DataFlow. The University also has an infrastructure project funded under the 2011-13 JISC MRD programme – DaMaRO – to develop a research data policy and roll out data management services across the institution.

The University of Southampton has long had strengths in terms of repository development. The role of repositories and data centres has been explored through the DataShare project and various crystallography projects, such as eCrystals and eBank. Like the university of Oxford, Southampton has been funded under both JISC MRD programmes to develop institutional infrastructure. The University developed a ten-year roadmap for research data management in the Institutional Data Management Blueprint (IDMB) project and is embedding this work through DataPool.⁵

Each of these examples demonstrates the need to approach the challenge holistically. Various strands of activity are needed, with input and collaboration across a variety of stakeholders. The 2011-13 JISC MRD programme is building on these early successful models: institutions funded to develop or embed data management infrastructure are also required to create high-level policies, guidance and associated support services.

⁴ Details of the various projects are available at: <http://www.ict.ox.ac.uk/odit/projects/datamanagement/>

⁵ IDMB project: <http://www.southamptondata.org/index.html>

What is Needed to Transform Policy into Practice?

It is clear that research data management is gaining significance at an institutional level, and researchers are increasingly developing data management and sharing plans to submit with grant proposals as a result of funders' policy requirements. However, these actions do not guarantee changes in practice. A diagram of stakeholders provided by Mark Thorley at the 7th Research Data Management Forum helps to illuminate this point.

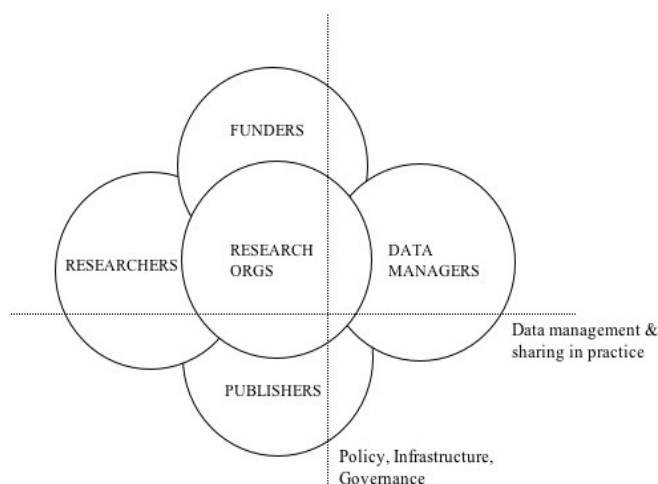


Figure 4. Stakeholders in research data management. Diagram inspired from version in Mark Thorley's presentation at the 7th Research Data Management Forum on Incentivising Data Management and Sharing. © 2011 DCC.

Policies and infrastructure fall in the realm of organisations such as research funders, universities and publishers, whereas actual data management activity is undertaken by stakeholders on the horizontal axis, namely researchers and data managers. Policies have most traction with political bodies, hence the marked response to the EPSRC policy. To move from policies into practice we need to flip to the other axis and focus on what motivates individuals.

Policies and infrastructure are needed, but it requires more than that to change practice. The dangers of simply introducing a mandate and accompanying infrastructure are well known.⁶ To encourage better practices across the board, infrastructure needs to be in line with user needs, and more importantly, rewards need to be in place. Recognition in terms of research assessment and career development would be powerful motivators. Data sharing occurs most coherently in cases where the community sees scientific benefit or simply has to share to undertake research, such as in molecular biology (ODE, 2011). Similarly, good data management practices often form out of negative experiences, such as data loss. Data management and sharing practice is down to individual will and self-interest, so to turn policies into practice we need to persuade people.

⁶ For example, see Dorothea Salo's presentation: "Le IR c'est mort" at <http://www.slideshare.net/cavlec/le-ir-cest-mort-vive-le-ir> and Melissa Terras' blog post: "On Thumb Twiddling" at <http://melissaterras.blogspot.com/2011/11/on-thumb-twiddling.html>

Conclusions

A policy is defined as a principle or prudent course of action. The term is not normally used to denote what is actually done, and therein lies the challenge. Few people do things because they 'should'. We act because we want to, because we believe in the cause, or see a direct benefit.

Policies alone will not inspire good practice; they are simply levers to motivate the people and processes that can enact change. The EPSRC policy demonstrates this, as assigning responsibility to institutions has provoked a marked response.

Just as good data management facilitates the goal of data sharing, policies are but one step in the process. The timeline shows how long it has taken to transform general principles into more practical policies. Institutions are now starting to respond by developing infrastructure, but we need to allow a similar transition period for these to be embedded and reward mechanisms to develop so policies can become practice.

Acknowledgements

This paper is based on work undertaken for the Digital Curation Centre and in support of the JISC Managing Research Data programmes. Thanks are extended to colleagues in both of these areas who have shared expertise.

References

- Arts and Humanities Research Council. (2011). *Research funding guide, version 1.7*. Retrieved from <http://www.ahrc.ac.uk/FundingOpportunities/Documents/Research%20Funding%20Guide.pdf>
- Biotechnology and Biological Sciences Research Council. (2010). *Data sharing policy*. Retrieved from <http://www.bbsrc.ac.uk/organisation/policies/position/policy/data-sharing-policy.aspx>
- Cancer Research UK. (2009). *Policy on data sharing and preservation*. Retrieved from <http://science.cancerresearchuk.org/funding/terms-conditions/funding-policies/policy-data-sharing/>
- Engineering and Physical Sciences Research Council. (2011). *Policy framework on research data*. Retrieved from <http://www.epsrc.ac.uk/about/standards/researchdata/Pages/default.aspx>
- Economic and Social Research Council. (2010). *Research data policy*. Retrieved from http://www.esrc.ac.uk/_images/Research_Data_Policy_2010_tcm8-4595.pdf
- Medical Research Council. (2011). *Policy on research data-sharing*. Retrieved from <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharing/Policy/index.htm>

- Natural Environment Research Council. (2010) *Data policy*. Retrieved from <http://www.nerc.ac.uk/research/sites/data/policy2011.asp>
- OECD. (2007). Principles and guidelines for access to research data from public funding. Retrieved from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- Opportunities for Data Exchange (ODE) Project. (2011). *Ten tales of drivers and barriers in data sharing*. Retrieved from www.stm-assoc.org/2011_10_22_ODE_brochure.pdf
- Research Councils UK. (1998). *Safeguarding good scientific practice*.
- Research Councils UK. (2011a). *Common principles on data policy*. Retrieved from <http://www.rcuk.ac.uk/research/Pages/DataPolicy.aspx>
- Research Councils UK. (2011b). *Policy and code of conduct on the governance of good research conduct: Integrity, clarity and good management*. Retrieved from <http://www.rcuk.ac.uk/documents/reviews/grc/goodresearchconductcode.pdf>
- Rice, R. & Haywood, J. (2011). Research data management initiatives at University of Edinburgh. *International Journal of Digital Curation*, 2(6). Retrieved from <http://www.ijdc.net/index.php/ijdc/article/view/194/259>
- Science and Technologies Facilities Council. (2011). *Scientific data policy*. Retrieved from http://www.stfc.ac.uk/Resources/pdf/STFC_Scientific_Data_Policy.pdf
- University of Edinburgh. (2011). *Research data management policy*. Retrieved from <http://www.ed.ac.uk/is/research-data-policy>
- UK Research Integrity Office. (2009). *Code of practice for research: promoting good practice and preventing misconduct*. Retrieved from <http://www.ukrio.org/what-we-do/code-of-practice-for-research/>
- Wellcome Trust. (2011). *Sharing research data to improve public health*. Retrieved from <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Data-sharing/Public-health-and-epidemiology/WTDV030690.htm>
- Wellcome Trust. (2010). *Policy on data management and sharing*. Retrieved from <http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm>